

This week:

- Rewrite patent :

本周的主要工作是把专利拆开分 3 个专利。

- 一种城市数据的跨域查询方法

这一部分主要介绍数据查询的基本方法,包括原子查询信息抽取和查询序列。前期的工作包括数据的属性分类。

- 一种城市数据的可视分析方法

这一部分主要讲可视化方法和可视化编码,主要是信息的可视化和可视化查询推理方法。

- 一种城市数据可视分析系统的实现方法

这一部分主要写的是系统的实现方法和系统实现流程。

- Write project introduction 智慧城市与可视分析

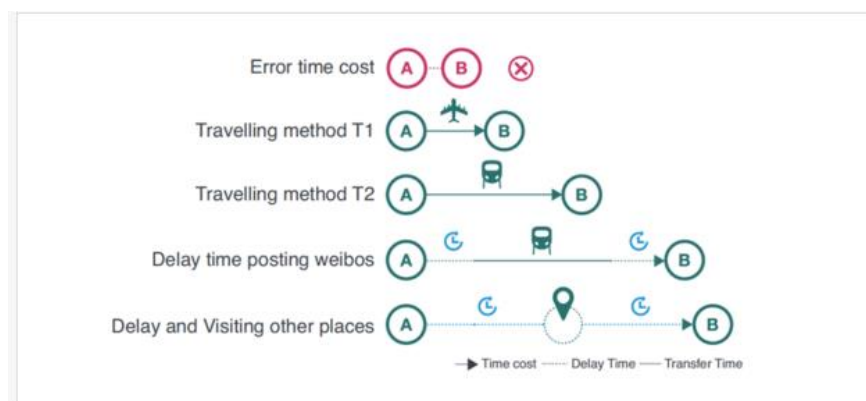
开始写了一个浮夸的: 智慧城市的狂潮已经蓄势待发,可视分析的暗涌已经悄然而至。数据挖掘,界面交互已经就位。如何利用城市数据管理城市,监控城市,如何利用数据成为城市的福尔摩斯,一切由你决定。大数据驱动未来,而你可以驱动大数据,智慧的城市期待智慧的你。

然而好像并不符合网上的规范,于是乎改成了项目介绍:智慧城市越来越成为当今时代的潮流。大量的城市数据需要数据专家和计算机专家的共同探索。城市数据的可视分析显得尤为重要。本项目中,我们希望通过城市数据的数据挖掘及其可视分析,挖掘城市数据中有价值的信息。系统需要完成城市数据的过滤查询,探索轨迹语意并对数据进行可视化。项目希望对数据不熟悉的普通用户也可以对系统进行操作并对数据进行探索。

- Study algorithm and programing with python.编程实现并撰写报告
 - EM：当有部分数据缺失或者无法观察到时，EM 算法提供了一个高效的迭代程序用来计算这些数据的最大似然估计。在每一步迭代分为两个步骤：期望（Expectation）步骤和最大化（Maximization）步骤，因此称为 EM 算法。
 - Levenberg-Marquardt：Levenberg-Marquardt 算法是最优化算法中的一种。最优化是寻找使得函数值最小的参数向量。它是使用最广泛的非线性最小二乘算法。
 - SVM：一种训练机器学习的算法，可以用于解决分类和回归问题。给定一些数据点，它们分别属于两个不同的类，现在要找到一个线性分类器把这些数据分成两类--这就是最基本的线性可分。
- Read paper 本周由于专利关系精读了一篇，但是这篇文章非常好。

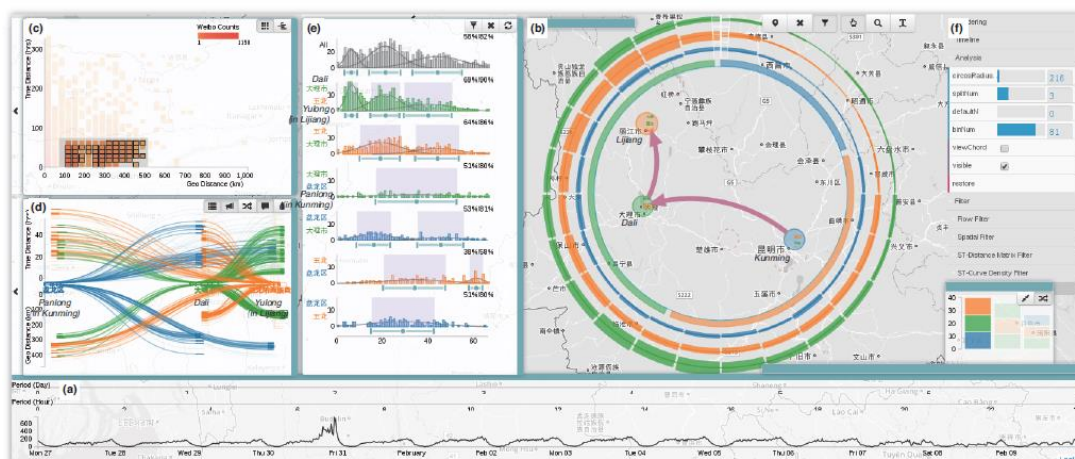
Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-tagged Social Media Data

这篇文章着实让我眼前一亮。确实是一篇非常好的工作。文章围绕微博数据展开，微博数据具有时间和空间属性，作者想要根据微博的时间空间属性来挖掘城市之间的移动时间和移动规律。例如下图从城市 a 移动到城市 b。然而一位微博数据的特殊性质，人的移动方式可能是不同的，坐火车坐飞机等。发微博的行为规律也是不同的，例如有人喜欢下飞机就发，有人则是玩了一天晚上再发微博。

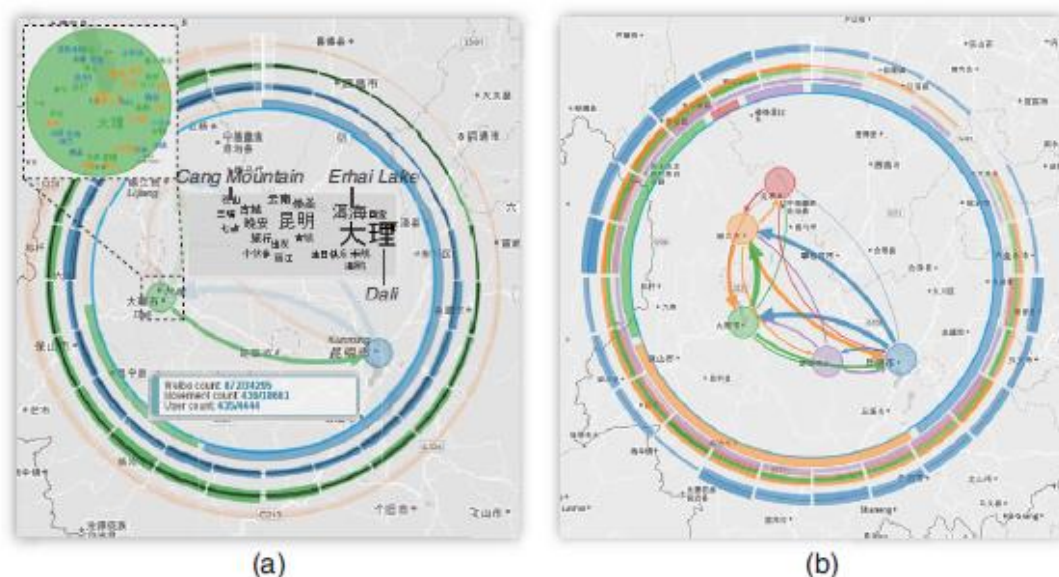


本文介绍了一种不是针对个体而是针对群体的，基于高斯混合概率模型挖掘方法的分析系统。这个系统能够帮助用户分析城市之间的移动时间，移动规律等。

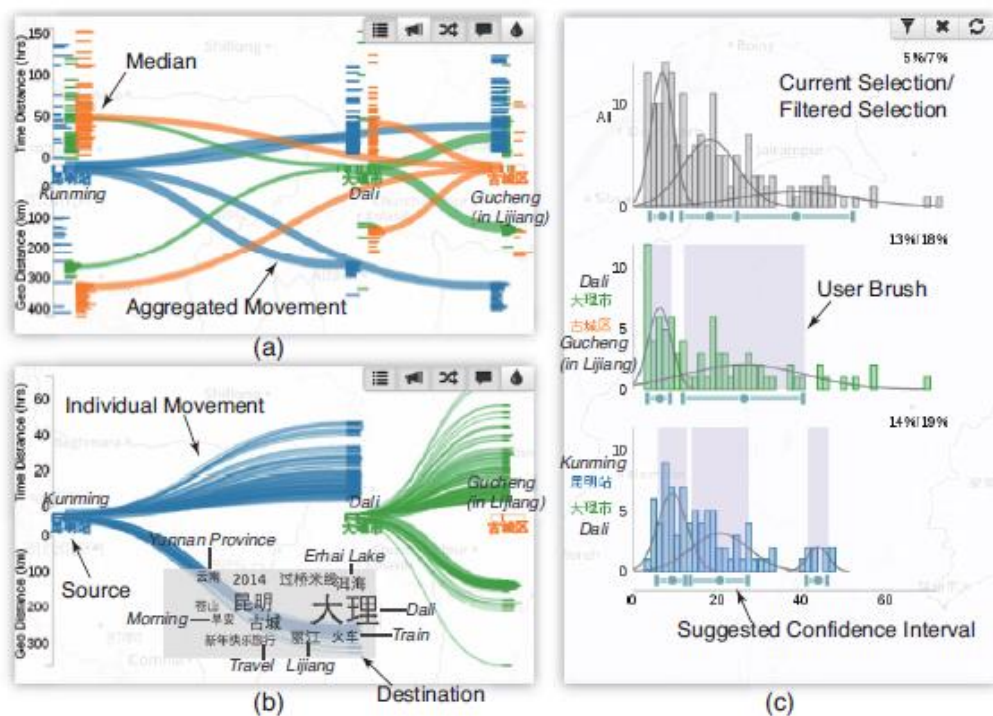
它的界面如下图



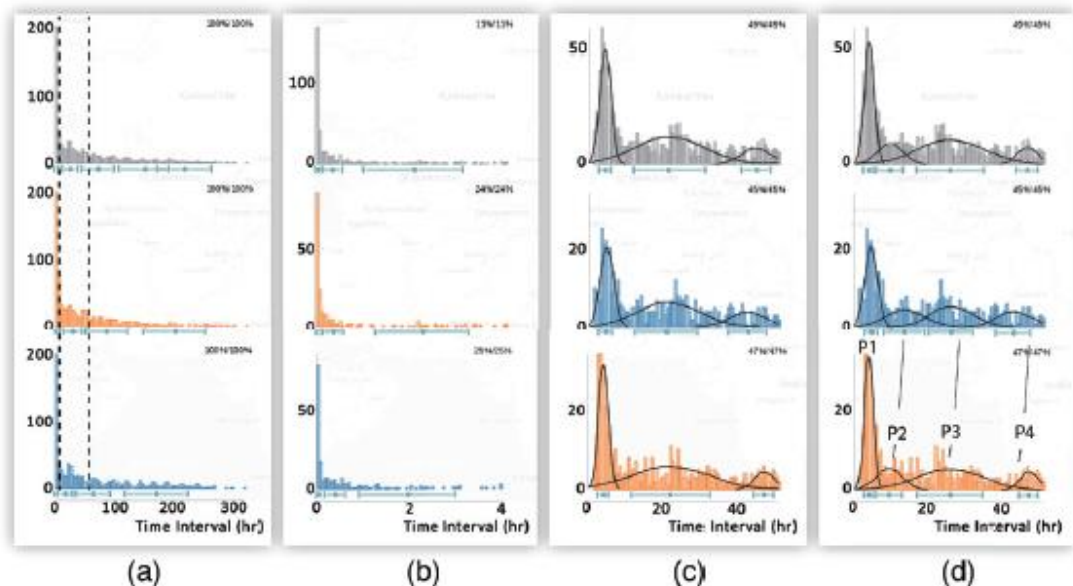
一个地图来展示空间，一个时间轴来展示当前分析的微博数据的时间分布。还有很多个详细分析视图来辅助用户进行分析。下面我还是记录一下他是怎么分析的：



在这个地图上，用户可以对微博数据进行简单的过滤。文章主要研究 OD 轨迹。所以用户可以指定起点终点和时间来过滤符合条件的微博。例如上图选择了大理昆明和丽江三个地点，那么三个城市之间移动的人如果在这三个不同的城市发布了微博，则会被过滤出来。外面的圆圈颜色对应地理上的位置颜色，圆环厚度表示了微博的数量。最外层还有一个圆圈对应所有微博的时间分布。在这些圆上，用户可以进一步对数据进行过滤。例如选择早上从大理到昆明的微博用户。地点之间的箭头粗细表示了符合移动条件的用户的数量。当鼠标悬浮时，系统高亮这些用户的详细信息和他们微博的词云统计。



然后详细信息的统计如上图。左边是对移动时间和移动距离的类似于平行坐标似的视图。右边的是重点，他采用了高斯混合概率模型来分析微博数据的移动时间统计数据，定好参数后找符合条件的符合正态分布的几个峰。它的具体操作流程如下：



首先进行统计，然后进行过滤异常数据，例如历时太短和太长的数据。然后进行控制参数进行估计。C 的参数为 3，发现三个峰，鼠标框选后发现第一个峰的微博词云显示飞

机较多，第二个峰为火车，第三个峰为延迟发微博的人。然后调整参数为 4，发现 p2 这个峰微博词云中显示晚点较多。

本文采用的方法给我的启示是用不确定性模型来分析分群移动轨迹。不是针对个人而是对人群移动进行分析，并采用高斯混合模型的不确定性建模来聚类，用户可以探索并实时的反馈进行新的探索。词云的应用也恰到好处。我们的数据比他广，所以我想我的上个星期的唯心主义的可视探索系统，也许也可以用一些不确定性模型来提示用户探索方向，并在展示数据轨迹的时候进行适当的聚类整合。

另外，本文的配色也很好看。文章中还有个流程图画的不错。值得学习。

Next week:

- Read more papers about Traj.
- Go home.